Articulated Shape Matching Using Locally Linear Embedding and Orthogonal Alignment

Diana Mateus, Fabio Cuzzolin, Radu Horaud, Edmond Boyer

Articulated Shape Matching Using Locally Linear Embedding and Orthogonal Alignment

Diana Mateus, Fabio Cuzzolin, Radu Horaud, and Edmond Boyer
INRIA Rhône-Alpes
655, avenue de l'Europe, Montbonnot Saint-Martin, FRANCE
name.latsname@inrialpes.fr

Abstract

In this paper we propose a method for matching articulated shapes represented as large sets of 3D points by aligning the corresponding embedded clouds generated by locally linear embedding. In particular we show that the problem is equivalent to aligning two sets of points under an orthogonal *transformation acting onto the* d*-dimensional embeddings. The method may well be viewed as belonging to the* model-based clustering *framework and is implemented as an EM algorithm that alternates between the estimation of correspondences between data-points and the estimation of an optimal alignment transformation. Correspondences are initialized by embedding one set of data-points onto the other one through out-of-sample extension. Results for pairs of voxelsets representing moving persons are presented. Empirical evidence on the influence of the dimension of the embedding space is provided, suggesting that working with higher-dimensional spaces helps matching in challenging real-world scenarios, without collateral effects on the convergence.*

1. Introduction

Shape matching is a central problem in computer vision as it allows to find shape classes for object recognition, to track objects over time, to build spatio-temporal representations useful for shape modeling, for action and/or gesture recognition, etc. Although methods are available both for rigid objects and deformable surfaces, articulated shape matching remains a challenging problem. Rigidity (and hence isometry) is only locally preserved and knowledge about how an articulated shape is split into rigid pieces is often not available. Whenever an object is represented by a cloud of 2D or 3D points, matching two different poses of the same object reduces to establishing assignments between points of the two clouds. Situations involving occlusions, missing data, outliers, and noise have already been addressed within the framework of matching *rigid* objects. A number of methods were suggested including hypothesize-and-test implemented as tree search, iterative closest point (ICP), and probabilistic assignment.

In addition to the difficulties associated with rigid alignment mentioned above, the problem of articulated alignment is more complex for at least two reasons: (i) Shape isometry is preserved only locally and not globally and (ii) object sub-parts may coalesce together, as is the case of a human with an arm lying along the torso. The lack of a global transformation that, in theory, maps points from one pose onto points of another pose, leads us to consider the more general problem of *maximum subgraph matching*.

As we claim here, those obstacles can be overcome by *aligning the two clouds of points in an embedding space*. The first method known by us to solve the weighted graph matching problem through an eigen (or spectral) decomposition was proposed by Umeyama [19]. More recently, Several spectral methods were proposed [1, 18] which compute non-linear embeddings of the input dataset by means of the SVD decomposition of an affinity matrix which depends on the structure of the data. ISOMAP, for instance, is based on the matrix of geodesic distances between points which are substantially preserved under articulated motion (factoring out changes in the topology of the moving body), as pointed out by other researchers [4]. By intuition, if the affinity matrix captures only local isometric properties of the input cloud, the shape of the resulting embedded cloud is only weakly affected by articulated deformations. In consequence, two clouds associated with different *articulated* poses of the same object can be *rigidly aligned in the embedding space*, such that each point of the first cloud is (in principle) associated with its nearest neighbor in the second (aligned) cloud.

In practice, as discussed here as well as in a companion paper [14], embedded shapes can only be aligned up to a $d \times d$ *orthogonal transformation*, where d is the dimensionality of the embedded space. Unlike its Euclidean sub-group, the orthogonal group does not have a Lie struc-

978-1-4244-1631-8/07/$25.00 ©2007 IEEE

ture. This means that the space of possible alignments under an orthogonal transformation is much larger than the space of alignments under rigid transformations, and closed-form solutions based on aligning the second order moments of the two sets are not available.

In this paper we propose to match articulated objects through *Locally Linear Embedding* (LLE) [16], in which the affinity matrix M does not possess an immediate interpretation in terms of pairwise distances between data-points. We show that the local isometry typical of LLE is enough to guarantee remarkable locally-rigid invariance under articulated motion. The results we obtain seem in fact to support the fact the LLE combines a good performance with an acceptable computational cost with respect to methods like ISOMAP.

We develop an EM method that alternates between point-to-point (or node-to-node) assignment and *orthogonal alignment* in the d-dimensional embedded space. Unlike other point-matching methods, we address the problem within the framework of *model-based clustering* [10]. In fact, each node of the smaller graph is viewed as a potential cluster with normal distribution, and there is an additional *outlier cluster* with uniform distribution [11]. The problem of matching then becomes the problem of *assigning* each node of the larger graph to one of these clusters. We also focus on the critical issue of how to initialize the alignment procedure. We exploit the mechanism of *out-of-sample extensions* [2] to determine an initial (probabilistic) guess of the correspondence function, by embedding points of one shape in the embedding space of the other one. Finally, we consider the crucial point of assessing the influence of the dimension of the embedding space d (the number of eigenvectors of M we select to compute the embedding). Even though some theoretical arguments suggest to choose d smaller than the number of zero eigenvalues of the matrix, we provide empirical evidence on the fact that in absence of noise higher dimension ensures better matching scores, while when dealing with real data there typically exists a threshold over which the performance of the algorithm degrades.

The remainder of the paper articulates as follows. In Section 2 we argue that local spectral embeddings (and LLE in particular) of articulated shapes are (reasonably) invariant under articulated motion. Associations between data-points can then be found by looking for nearest neighbors after alignment. This invariance is tempered by inherent ambiguities which are solved by means of an alternating EM estimation of matching and optimal alignment (Section 3). We show how to initialize the procedure by means of out-of-sample extension, and discuss the influence of the parameter d. Finally (Section 4) we present results for both simplified tests in which ground truth is available and real-world data.

2. Matching in the embedding space

Pose-invariance of embedded clouds under articulated motion. LLE [16] is an unsupervised learning algorithm which computes embeddings $\{\mathbf{x}_i\}$ of a set of input points $\{\mathbf{X}_i\}$, $i = 1, ..., n$, while preserving the local structure of the data, i.e., the distances between each point and its k neighbors, k being a parameter of the algorithm. These optimal embeddings (up to a global rotation of the whole space) are found by selecting the eigenvectors associated with the bottom $d + 1$ eigenvalues of the affinity matrix $M_{ij} \doteq \delta_{ij} - W_{ij} - W_{ji} + \sum_k W_{ki} W_{kj}$ ($\delta_{ij} = 1$ iff $i = j$, 0 otherwise), where W_{ij}, $j = 1, ..., k$ are the weights that best linearly reconstruct $\mathbf{X}_i$ from its neighbors: $\arg\min_W \|\mathbf{X}_i - \sum_j W_{ij}\mathbf{X}_j\|^2$. The embedded cloud is constrained to be centered at the origin $\sum_i \mathbf{x}_i = \mathbf{0}$, and have unit covariance.

As it has been noticed before, some embedding schemes (like ISOMAP) are inherently *pose-invariant* under articulated motion (since geodesic distances between pairs of points do not change as the kinematic motion of the articulated object proceeds). This is not true, in a strict sense, for LLE. However, as its embedding depends only on the local structure of the input dataset, under articulated motion the shape of the LLE embedded cloud exhibits indeed remarkable stability. All local neighborhoods incident on a rigid part of an articulated body are preserved along the motion, while only the few neighborhoods interested by evolving joint(s) are affected (Figure 1-left). The affinity matrix M undergoes only little changes, depending of course on the values of n and k (as smaller ks reduce the number of neighborhoods with non-empty intersection with moving joints) and the number of evolving joints. Figure 1 shows some anecdotal evidence supporting this claim. Similar considerations hold for Laplacian Eigenmaps [1].

Intrinsic ambiguities in eigenspace alignment. Matching shapes in the embedding space is, in principle, only weakly affected by articulated deformations. It is therefore tempting to conjecture that shape matching is now reduced to eigenspace alignment [19], [9], [12].

Nevertheless, several issues emerge:

- eigenvectors are only defined up to sign, which introduces 2^d possible alignments between the two eigenbases (*sign-reversal ambiguity*);
- the ordering of eigenvalues is not reliable due to eigenvalues with large algebraic multiplicity and/or numerical instabilities in their calculation. In the work of Umeyama [19] this was not an issue since the author considered very small graphs. In recent work, the problem of eigenvalue ordering has been overlooked. We have then to consider $d!$ possible permutations between the eigenvalues (*eigenvalue-ordering ambiguity*).
- LLE embedding is defined up to a *residual* transformation ($d \times d$ rotation) between the two eigenbases (*rigid-motion*

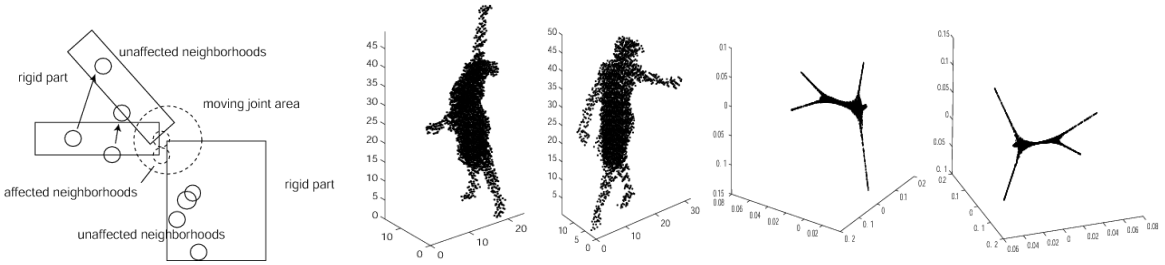

Figure 1. Left: The number of neighborhoods affected by articulated motion is relatively small. Right: Some anecdotal evidence on the stability of locally linear embedding under articulated motion. Different poses of the same articulated body are mapped to the same embedded cloud, for a large interval of parameter values (here $d = 3$, $k = 10$).

ambiguity). - perfect match between two embedded clouds is unlikely, due to spurious points, missing, bad, and/or noisy data, etc.

As we argue in the following, these issues can be overcome by searching for an optimal *orthogonal transformation* Q that allows the alignment of the largest number of point-to-point (or node-to-node) assignments.

3. Methodology

Let us denote by $\mathbf{X} = \{\mathbf{X}_i\}_{1\leq i\leq n}$ and by $\mathbf{Y} = \{\mathbf{Y}_j\}_{1\leq j\leq m}$ the two sets of points, and with $\{\mathbf{x}_i\}_{1\leq i\leq n}$ and $\{\mathbf{y}_i\}_{1\leq i\leq m}$ the corresponding sets of d-dimensional embedded points generated through LLE (Section 2). According to what has been discussed above, the $d\times d$ matrix Q that allows, in principle, to align the two embedded clouds has the form $Q = RPS$, where R is a $d\times d$ rotation, P is a $d\times d$ permutation matrix, and S is a $d \times d$ diagonal matrix with entries $s_{ii} = \pm 1$ encoding rotational, eigenvalue-ordering, and sign-reversal ambiguities respectively. As R, P, and S are orthogonal, Q is an orthogonal matrix as well. Points in the two sets can now be aligned according to $\mathbf{x}_{z(j)} = Q\mathbf{y}_j$, with $z : \mathbf{x} \rightarrow \mathbf{y}$ the correspondence function. Notice that the mapping $\mathbf{x} \rightarrow \mathbf{y}$ directly implies the mapping $\mathbf{X} \rightarrow \mathbf{Y}$. As both z and Q are unknown, the embedded alignment can be cast into the following minimization problem:

$$\min_{Q,z} \sum_j \|\mathbf{x}_{z(j)} - Q\mathbf{y}_j\|^2 \qquad (1)$$

This formulation is analoguous with rigid alignment [22] where the $d \times d$ orthogonal matrix is reduced to a 4×4 rigid transformation. A similar approach was proposed in [8] for articulated tracking in kinematic space.

An EM framework for aligning embedded spaces. We can pose the problem in a probabilistic framework by defining $\mathbf{x} = \{\mathbf{x}_i\}_{1\leq i\leq n}$ as a set of observed values of an equal number of random variables that will be denoted by $\mathcal{X} = \{\mathcal{X}_i\}$. To each random variable $\mathcal{X}_i$ we associate another random variable z_i which describes the correspondence. Specifically, $z_i = \mathbf{y}_j$ will mean that the *observation* $\mathbf{x}_i$ is in correspondence with the *predicted* model point $Q\mathbf{y}_j$, while $z_i = \emptyset$ means that observation $\mathbf{x}_i$ belongs to an *outlier cluster*. Let v be the volume within which $\mathbf{y}_j$ is to be expected. Without loss of generality, this volume is a sphere with radious σ_0, $v = 3\pi\sigma_0^3/4$. The prior probability that an observation point $\mathbf{x}_i$ is assigned to a model point $\mathbf{y}_j$ writes $P(z_i = \mathbf{y}_j) = v/V$, where V is the volume of the whole space. In order to satisfy $\sum_j P(z_i = \mathbf{y}_j) + P(z_i = \emptyset) = 1$, we set $P(z_i = \emptyset) = (V - mv)/V$. The main difference between our approach and similar ones [5, 13] is that the two sets of points are not treated symmetrically [21], [8]. We choose to model with different distributions the inliers and the outliers, [21], [11]. The probability of an observation $\mathbf{x}_i$ to lie in the proximity of $Q\mathbf{y}_j$, given that $\mathbf{x}_i$ and $\mathbf{y}_j$ are in correspondence, will be described by a Gaussian distribution with covariance matrix σI_d (f denotes the Euclidean distance):

$$P_Q(\mathbf{x}_i|z_i = \mathbf{y}_j) = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\frac{f^2(\mathbf{x}_i, Q\mathbf{y}_j)}{2\sigma^2}}. \qquad (2)$$

The probability of an observation given that it corresponds to an outlier will be described by a uniform distribution over the volume of the working space: $P(\mathbf{x}_i|z_i = \emptyset) = 1/V$. We look for the posterior probability that $\mathbf{x}_i$ corresponds to $\mathbf{y}_j$ given the observation $\mathbf{x}_i$:

$$\alpha_{ij} \doteq P_Q(z_i = \mathbf{y}_j|\mathbf{x}_i).$$

The above expressions for priors and likelihoods can be combined in the Bayesian framework (after observing that the set $\{z_i = \mathbf{y}_1, \ldots, z_i = \mathbf{y}_m, z_i = \emptyset\}$ is a partition of the event space) as $P_Q(\mathbf{x}_i) = \sum_{j=1}^m P(\mathbf{x}_i|z_i = \mathbf{y}_j)P(z_i = \mathbf{y}_j) + P(\mathbf{x}_i|z_i = \emptyset)P(z_i = \emptyset)$ so that (with a proper choice for σ_0):

$$\alpha_{ij} = \Big(e^{-\frac{f^2(\mathbf{x}_i, Q\mathbf{y}_j)}{2\sigma^2}}\Big) \Big/ \Big(\sum_j e^{-\frac{f^2(\mathbf{x}_i, Q\mathbf{y}_j)}{2\sigma^2}} + \sigma^3\Big). \qquad (3)$$

In order to estimate the transformation Q we maximize the expectation E of the logarithm of the joint probability of the observations and their correspondences $F(Q, Q^c) =$

$E_{Q^c}[\log(P_Q(\mathbf{x},z))|\mathbf{x}]$ where Q^c denotes the current estimate of Q. Assuming $\mathbf{x}_i$ and z_i independent $\forall\ 1 \leq i \leq n$ we get

$$P_Q(\mathbf{x},z) = \prod_{i=1}^{n}\Big(\prod_{j=1}^{m}(P(\mathbf{x}_i|z_i=\mathbf{y}_j)P(z_i=\mathbf{y}_j))^{\delta_{y_j}(z_i)} \\ (P(\mathbf{x}_i|z_i=\emptyset)P(z_i=\emptyset))^{\delta_{\emptyset}(z_i)}\Big)$$

where $\delta_{y_j}(z_i)$ (respectively $\delta_{\emptyset}(z_i)$) equals 1 when $z_i = \mathbf{y}_j$ (respectively $z_i = \emptyset$), and 0 otherwise. After some straightforward algebraic manipulations and by grouping constant terms, we obtain:

$$F(Q,Q^c) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_{ij}(\|\mathbf{x}_i - Q\mathbf{y}_j\|^2 + a\log\sigma). \quad (4)$$

We can further simplify this expression by replacing $\mathbf{w}_j = \sum_i \alpha_{ij}\mathbf{x}_i / \sum_i \alpha_{ij}$ and $\beta_j = \sum_i \alpha_{ij}$ in (4). Maximizing $F(Q,Q^c)$ is then equivalent to:

$$\min_{Q}\sum_{j=1}^{m}\beta_j(\|\mathbf{w}_j - Q\mathbf{y}_j\|^2 + a\log\sigma). \quad (5)$$

Now, the probabilities α_{ij} can be treated as hidden variables within the Expectation-Maximization framework [7], in which Q and σ on one side, and the α_{ij}'s on the ather side, are estimated. The assignment probabilities are computed in the E step, while in the M step Q and σ are estimated by solving (5). Notice that the estimation of Q amounts to find the mean $Q\mathbf{y}_j$ of each Gaussian cluster j. Q can be computed in closed-form by relaxing the rotation constraint in [20]. Namely, we compute the singular value decomposition $H = U\Lambda V^\top$ of the matrix $H = \sum_j \beta_j \mathbf{w}_j \mathbf{y}_j^\top$ and retain the solution $Q = VU^\top$.

Matching initialization through out-of-sample extensions. The above EM estimation can be initialized either in the M-step (by finding an initial value for the orthogonal matrix Q) or in the E-step, by choosing an initial assignment matrix $\bar{\alpha} = [\alpha_{ij}]$. Here we propose a natural way to initialize $\bar{\alpha}$ by means of *out-of-sample extensions* of spectral embeddings to points *not* in the training set. In [2], the extension to new points modifies the embedding for all original training points, fact which violates our principle of aligning embedded shapes independently generated. In [17], instead, the addition of a point outside the training set does not affect the affinity matrix M on which the embedding is based, as the new point is not considered part of the neighborhood of any training point. In opposition, its own neighborhood is detected in the usual manner as a set of k training points. The corresponding weights are computed and then used to reconstruct the location of the point in the existing embedded space.

Assume now that the two shapes correspond to two poses of an articulated shape which differ only in the position/configuration of a few limbs. In that case, most of the shape remains stable. Under these assumptions, out-of-sample embeddings $\mathbf{y}'_i$ of points $\mathbf{Y}_i$ of the second shape (Figure 2) which belong to the region shared with the first shape will be located close to their counterparts in the first embedded cloud. We can then assign to each point $\mathbf{x}_j$ of the first embedded cloud a matching likelihood proportional to its distance from $\mathbf{y}'$: $\alpha_{ij} \propto f(\mathbf{y}'_i, \mathbf{x}_j)$. The initialization

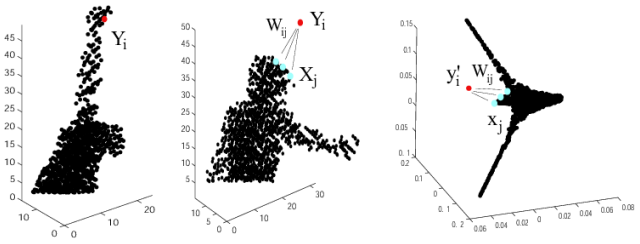

Figure 2. Out-of-sample embeddings (right) of points of the second shape (left) with respect to the first shape (middle).

algorithm reads then as follows: Given the embedded cloud $\{\mathbf{x}_i, i = 1, ..., n\}$ for the first shape:
1. for each point $\mathbf{Y}_i$ of the (original) second shape, detect its k neighbors *in the first shape*, and compute the related weights: $\mathbf{Y}_i = \sum_{j=1}^{k} W_{ij}\mathbf{X}_j$;
2. as LLE preserves local neighborhoods, those same weights can be used to find the out-of-sample embedding $\mathbf{y}'_i$ of $\mathbf{Y}_i$ in the embedding space *of the first shape*: $\mathbf{y}'_i = \sum_{j=1}^{k} W_{ij}\mathbf{x}_j$;
3. an initial estimate for the correspondences α_{ij} can then be computed by assessing the likelihood of those points with respect to a Gaussian distribution centered in the out-of-sample embedding $\mathbf{y}'_i$ of $\mathbf{Y}_i$: $\alpha_{ij} = \mathcal{N}(f(\mathbf{y}'_i, \mathbf{x}_j))$, where $f(\mathbf{y}'_i, \mathbf{x}_j)$ is the distance between $\mathbf{y}'_i$ and $\mathbf{x}_j$, the standard deviation σ being a parameter of the EM algorithm (Equation 3).
It is worth to stress that out-of-sample embeddings are only used to compute an initial guess of the association matrix, which is later used to align the two *independently* computed embedded clouds in the M-step.

Dimension of the embedding space. A crucial issue with LLE (and spectral methods in general) is the choice of the dimensionality d of the embedding space, i.e. the number of eigenvectors of the affinity matrix M. This issue has been in fact addressed, for instance by Polito and Perona [15]. They argue that the covariance constraint $\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \otimes \mathbf{x}_i = I$ causes an overfitting in the embedding when choosing the "wrong" dimension, and prove that any full-rank embedding associated with a zero-error approximation $\mathbf{x}_i = \sum_j W_{ij}\mathbf{X}_i$ has to be associated with $d < a$ where a is the number of zero eigenvalues of the affinity

matrix M.
In practice, though, we do not observe this effect. The reason is due in part to the inclusion of the unit-covariance constraint in the optimization problem which forces the numerically obtained solution not to meet the locally linear reconstruction constraint exactly. The other factor is the inclusion of a regularization term [17] in the computation of the covariance matrix which is often necessary when dealing with d greater than the original dimensionality of the point cloud: Again, this keeps the final solution (embedding) away from the ideal zero-error approximation.

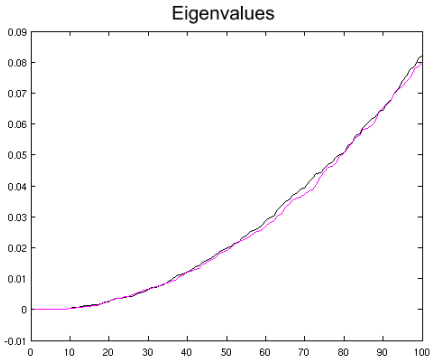

Figure 3. Typical behavior of the eigenvalues of the affinity matrix of a cloud of points.

Figure 3 shows a typical spectrum of M for a dataset $\mathbf{X}$ formed by 1,300 3D points representing a moving person. It can be noticed that there is neither clear evidence of what is the cardinality of the set of "zero" eigenvalues, nor a reliable way to order these eigenvalues, and as a consequence which value of d we should use.
In the context of clustering [6], on the other side, it has been pointed out that the number of dimensions to select has to match the desired number of clusters. We will show some empirical evidence on the influence of d in the data association problem in the last part of the paper.

Number of neighbors k. A parameter which influences the stability of embedded cloud (and hence the hypothesis of pose invariance on which the entire alignment scheme is based) is the number of neighbors k of the LLE algorithm. In [6] a method to tune the value of k has been proposed which relies on the detection of "anomalous" neighborhoods, i.e. neighborhoods which span separate body-parts (Figure 4-b-middle). These are characterized by the fact that their farthest elements (as they belong to another, distinct body-part) are relatively distant from all others (which instead lie all on the same rigid part). If we then plot the distance between the farthest point of the neighborhood and all its fellows we can notice a significant jump (Figure 4-b-bottom:right). This is not the case for neighborhoods which span a single rigid part (Figure 4-b-top:right). It is rather

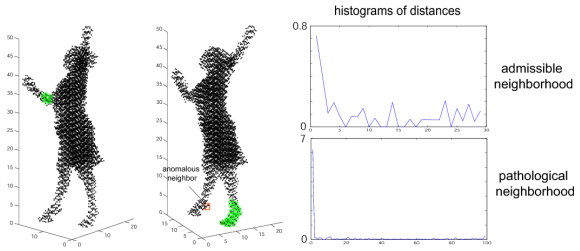

Figure 4. How to estimate the correct number of neighbors k in the LLE algorithm. Non admissible values of k are characterized by "pathological" neighborhoods (in green) which span distinct body-parts (middle), in opposition to admissible values (left). The corresponding distance plots are visible to the right.

natural to choose as correct k any of those values which yield only "regular" neighborhoods.

Summary of the proposed matching algorithm.
1. (*Initialization*) Initial values for the assignment matrix $\bar{\alpha}$ are set through out-of-sample extension; then alternate between:
2. (*Maximization*) Use Eq. (5) to find an estimate for Q. Allow the variance to decrease geometrically with $\sigma^{\text{new}} = \kappa\sigma$, $0 < \kappa < 1$;
3. (*Expectation*) The probabilities α_{ij} of each possible association (and consequently β_j and $\mathbf{w}_j$) are evaluated by computing (3) the Euclidean distance between each $\mathbf{x}_i$ and each $Q\mathbf{y}_j$ [3];
4. (*Test*) If the entries of Q are stabilized or if $\sigma^{\text{new}} \leq \sigma_{\min}$ then terminate, else go to step 2;
5. (*MAP*) $z_i = \arg\max_{j,\emptyset}\{\alpha_{ij}, \alpha_{i\emptyset}\}$.

4. Results

We tested our approach to data association in the context of articulated object motion. We acquired several image sequences using an acquisition system formed by 8 synchronized cameras, e.g., Figure 5. Silhouettes from all viewpoints were processed to compute their visual hull. The moving 3D articulated body (a person) was finally rendered as a uniformly sampled voxelset.

Tests on point permutations in rigid shapes. In a first series of experiments, though, we tested the coherence of the presented EM scheme by measuring its performance when matching rigid shapes formed by a large number of points. We applied random permutations to the ordering of the points of a given voxelset, and ran the algorithm of Section 3 to estimate the associations. We can compare them with the enforced ground truth and measure performances in terms of percentage of points for which the correct matching is associated with an α_{ij} among the largest c. We first applied 100 random permutations to a cloud of 1300 data-points to get a second cloud in absence of noise.

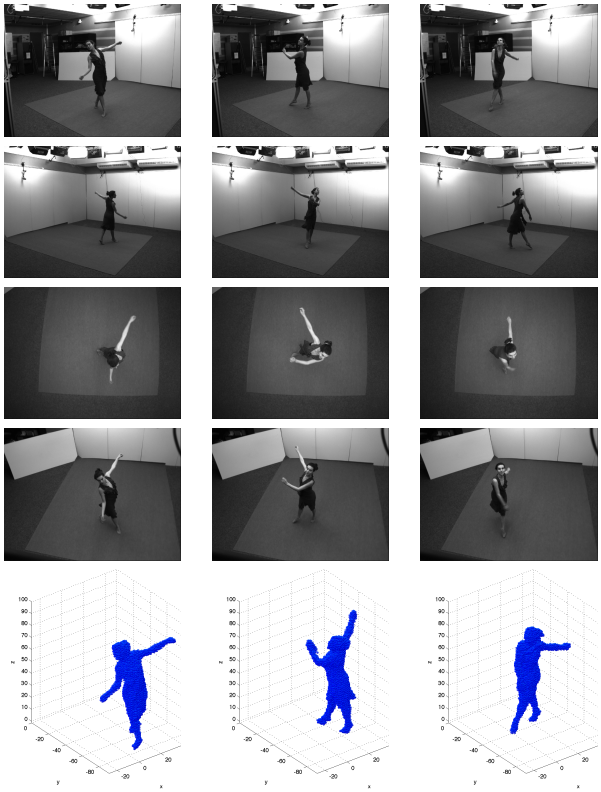

Figure 5. From top to bottom: Images associated with cameras #1, #6, #7, and #8 and the corresponding voxel set. From left to right: three different poses of a dancer.

Figure 6-top plots the percentage of *correct associations* ($c = 1$) for increasing values of d. We computed the score for 20 repeated trials, so that the diagram plots the average score together with the associated standard deviation. Our EM estimation correctly recovers the imposed permutations and matches the two datasets. We can also observe that the higher the value of d the better the matching rate, for different values of the LLE regularization term.

We later applied additive Gaussian noise (with variance comparable to the voxelgrid size) to each data-point after permuting as before a random subset of points to generate the second embedded cloud. Figure 6-middle plots the related matching score for $c = k = 13$. The method exhibits remarkable performances even in the presence of noise. It is quite interesting to notice that in this case the score has a maximum for $40 \leq d \leq 50$ and degrades for higher dimensional embeddings.

We also worked out a score which measures the local isometry of the two shapes after matching. If pose-invariance holds (ideal case) local neighborhoods of all points are preserved under optimal matching. If the estimated correspondence z is good the neighborhood of each point $\mathbf{X}_i$ of the first shape is mapped to the "same" (preserved) neighborhood in the second shape. Analytically, $iso(z) = \max_{i=1,\dots,n} \max_{j \in N(i)} |f(\mathbf{X}_i, \mathbf{X}_j) - f(\mathbf{Y}_{z(i)}, \mathbf{Y}_{z(j)})|$ measures over all the neighborhoods of the first shape, the maximum change of distance between the central point and its k-neighbors, when the mapping z is applied. Figure 6-bottom plots $iso(z)$ for the same experiment.

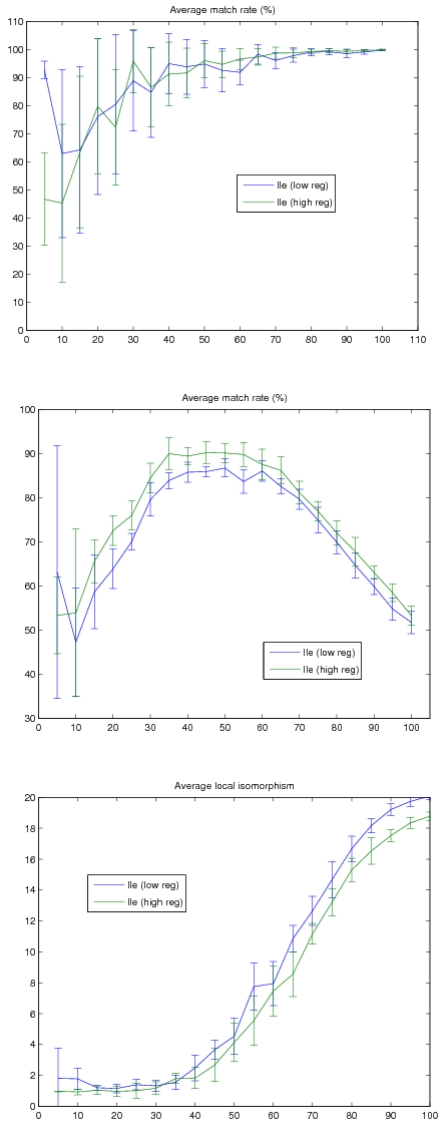

Figure 6. Matching score of the EM algorithm as a function of the dimension d of the embedding space. Lower and higher values of the regularization factor are compared. Top: In absence of noise (only permutations of the input data-points). Middle: When applying additive Gaussian noise to generate the second cloud. Average and standard deviation over 20 repeated trials are shown. Bottom: Corresponding $iso(z)$: an clear inverse correlation with the matching score can be noticed.

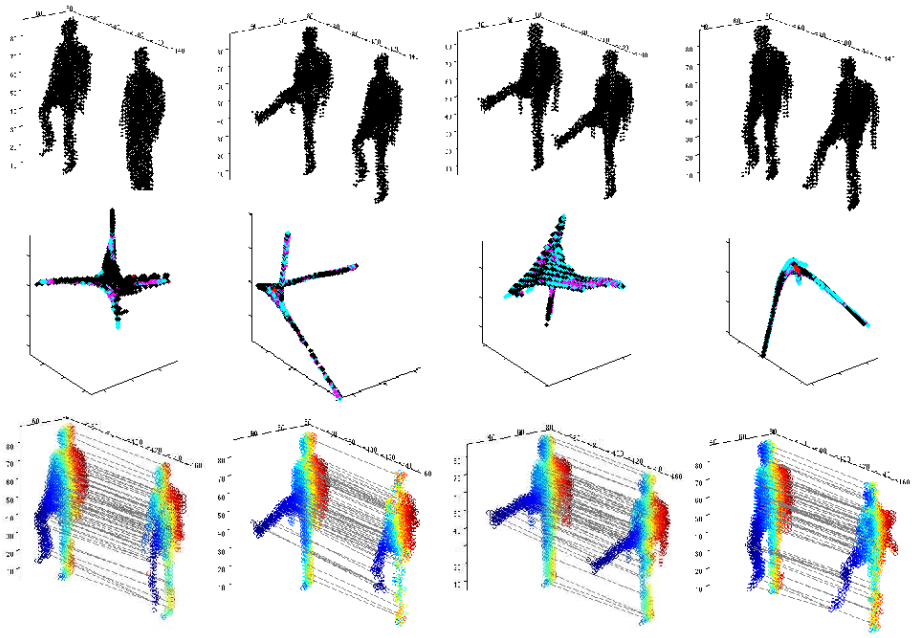

Figure 7. Matching results for pairs of poses coming from a real sequence of voxelsets (a person kicking the air). Top: Four pairs of poses extracted from the sequence. Middle: Optimal alignments of the corresponding embedded shapes (d = 15 but only the first three dimensions are shown). All pairs of embedded clouds (in black and color) are perfectly superimposed. Bottom: The related optimal correspondences are rendered as similar colors for corresponding parts of the body (a number of straight lines representing correspondences between a random selection of points are also plot).

Tests with real sequences under constrained motion. In a second series of tests we used sequences of voxelsets generated by the multi-camera system. Figure 7 pictorially shows typical matching results obtained for a number of sample poses coming from a sequence of a person kicking the air. For all pairs of clouds most of the points remain in roughly the same location, allowing the initialization through out-of-sample extension to work pretty well. The results in terms of matching (remember we deal with 1000-2000 points here) are quite impressive.

Dimension of the embedding space. Finally, we analyzed the influence of the embedding dimension parameter on matching performances in the case of real data. We plotted $iso(z)$ for pairs of poses coming from a sequence of voxelsets in which a person marches in a rather smooth way (allowing initialization based on out-of-sample extension to work). Figure 8 shows that, unlike the simplest case of pure permutation between data-points, there exist an optimal value of d (the number of selected eigenvectors) which correspond to the inflection point in the plots and varies according to different relative configurations of the pair of poses shown on Figure 7. Notice that no such thing is visible in the typical diagram of the eigenvalues of M reported in Figure 4.

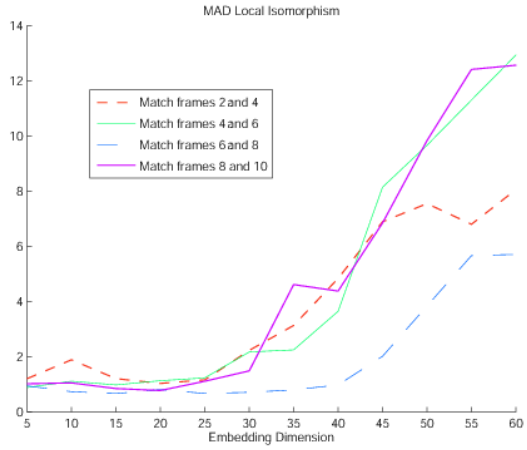

Figure 8. Matching EM-LLE performances (measured by the isometry score) plotted versus d for several different pairs of shapes extracted from a real sequence of voxelsets (Figure 7).

5. Perspectives

In conclusion, we presented an an alternating EM method, within the framework of model-based clustering, which matches articulated shapes composed by a large number of points through alignment in an embedding space.

We focused in particular on LLE which seems to combine good performance with an acceptable computational cost. Potential applications of such a method are enormous, and range from graph matching to shape recognition (as a matching score can be used as a measurement of the similarity of two shapes), to unsupervised (i.e. without knowledge of an underlying dynamical model) articulated tracking, to clustering of point trajectories (generated by associating points along time) for segmentation of body-parts or actions to be later classified. Concerning the dimension of the embedding, empirical evidence suggests that working with real data is qualitatively different from proving statements on the ideal case, and that the higher d the better (at least for matching). Increasing the dimension, in any case, does not affect the number of EM iterations. Last but not least, even though the method has been tested for articulated motion only it is reasonable to conjecture that it could be extended (under precise assumptions) to some classes of deformable bodies.

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, volume 14. 2002.

[2] Y. Bengio, J.-F. Paiement, and P. Vincent. Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering. Technical report, Departement d'Informatique et Recherche Operationelle, Universite' de Montreal, 2003.

[3] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] C.-W. Chu, O. C. Jenkins, and M. J. Mataric. Markerless kinematic model and motion capture from volume sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, pages 475–482, 2003.

[5] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, February 2003.

[6] F. Cuzzolin, D. Mateus, E. Boyer, and R. Horaud. Robust spectral 3D-bodypart segmentation along time. In *submitted to ICCV'07 2nd Workshop on HUMAN MOTION: Understanding, Modeling, Capture and Animation*, 2007.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[8] G. Dewaele, F. Devernay, R. Horaud, and F. Forbes. The alignment between 3-D data and articulated shapes with bending surfaces. In *Proceedings of ECCV'06*, pages 578–591. Springer, May 2006.

[9] A. Elad and R. Kimmel. On bending invariant signatures for surfaces. *Pattern Analysis and Machine Intelligence PAMI*, 25(10):1285–1295, 2003.

[10] C. Fraley and A. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97:611–631, 2002.

[11] C. Hennig and P. Coretto. The noise component in model-based cluster analysis. In *To appear in Proceedings of GfKl*, Freiburg, Germany, 2007.

[12] V. Jain and H. Zhang. Robust 3-d shape correspondence in the spectral domain. In *IEEE International Conference on Shape Modeling and Applications 2006*, 2006.

[13] B. Luo and E. Hancock. A unified framework for alignment and correspondence. *CVIU*, 92(1):26–55, October 2003.

[14] D. Mateus, F. Curzzolin, R. Horaud, and E. Boyer. Articulated shape matching by robust alignment of their embedded representations. In *Submitted to ICCV'07 Workshop on 3D Representation for Recognition (3dRR-07)*, 2007.

[15] M. Polito and P. Perona. Grouping and dimensionality reduction by locally linear embedding, 2001.

[16] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[17] L. K. Saul and S. T. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J. Mach. Learn. Res.*, 4:119–155, 2003.

[18] J. B. Tenenbaum, V. d. Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000.

[19] S. Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, May 1988.

[20] S. Umeyama. Least-squares estimation of transformation parameters between two point pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, April 1991.

[21] W. Wells III. Statistical approaches to feature-based object recognition. *International Journal of Computer Vision*, 28(1/2):63–98, 1997.

[22] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal on Computer Vision*, 13:119–152, 1994.